

#4

SEQUENCE LISTING

(1) GENERAL INFORMATION:

sub
C

(i) APPLICANT: Zavada, Jan
Pastorekova, Silvia
Pastorek, Jaromir

(ii) TITLE OF INVENTION: MN Gene and Protein

(iii) NUMBER OF SEQUENCES: 86

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Leona L. Lauder
(B) STREET: 369 Pine Street
(C) CITY: San Francisco
(D) STATE: California
(E) COUNTRY: USA
(F) ZIP: 94104

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US 09/772,719
(B) FILING DATE: 01-30-2001
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 08/485,049
(B) FILING DATE: 07-JUN-1995

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Lauder, Leona L.
(B) REGISTRATION NUMBER: 30,863
(C) REFERENCE/DOCKET NUMBER: D-0021.3A-2

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 415-981-2034
(B) TELEFAX: 415-981-0332

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1522 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

| | |
|--|-----|
| ACAGTCAGCC GCATGGCTCC CCTGTGCCCC AGCCCCTGGC TCCCTCTGTT GATCCCGGCC | 60 |
| CCTGCTCCAG GCCTCACTGT GCAACTGCTG CTGTCACTGC TGCTTCTGAT GCCTGTCCAT | 120 |
| CCCCAGAGGT TGCCCCGGAT GCAGGAGGAT TCCCCCTTGG GAGGAGGCTC TTCTGGGGAA | 180 |
| GATGACCCAC TGGGCGAGGA GGATCTGCCC AGTGAAGAGG ATTCACCCAG AGAGGAGGAT | 240 |
| CCACCCGGAG AGGAGGATCT ACCTGGAGAG GAGGATCTAC CTGGAGAGGA GGATCTACCT | 300 |
| GAAGTTAAGC CTAAATCAGA AGAAGAGGGC TCCCTGAAGT TAGAGGATCT ACCTACTGTT | 360 |
| GAGGCTCCTG GAGATCCTCA AGAACCCAG AATAATGCCC ACAGGGACAA AGAAGGGGAT | 420 |
| GACCAGAGTC ATTGGCGCTA TGGAGGCGAC CCGCCCTGGC CCCGGGTGTC CCCAGCCTGC | 480 |
| GCGGGCCGCT TCCAGTCCCC GGTGGATATC CGCCCCCAGC TCGCCGCCTT CTGCCCCGGCC | 540 |
| CTGCGCCCCC TGGAATCCT GGGCTTCCAG CTCCCGCCGC TCCAGAACT GCGCCTGCGC | 600 |
| AACAATGGCC ACAGTGTGCA ACTGACCCTG CCTCCTGGGC TAGAGATGGC TCTGGGTCCC | 660 |
| GGGCGGGAGT ACCGGGCTCT GCAGCTGCAT CTGCACTGGG GGGCTGCAGG TCGTCCGGGC | 720 |
| TCGGAGCACA CTGTGGAAGG CCACCGTTTC CCTGCCGAGA TCCACGTGGT TCACCTCAGC | 780 |

| | |
|---|------|
| ACCGCCTTTG CCAGAGTTGA CGAGGCCTTG GGGCGCCCGG GAGGCCTGGC CGTGTTGGCC | 840 |
| GCCTTTCTTG AGGAGGGCCC GGAAGAAAAC AGTGCCTATG AGCAGTTGCT GTCTCGCTTG | 900 |
| GAAGAAATCG CTGAGGAAGG CTCAGAGACT CAGGTCCCAG GACTGGACAT ATCTGCACTC | 960 |
| CTGCCCTCTG ACTTCAGCCG CTA CTCTCCAA TATGAGGGGT CTCTGACTAC ACCGCCCTGT | 1020 |
| GCCCAGGGTG TCATCTGGAC TGTGTTTAAC CAGACAGTGA TGCTGAGTGC TAAGCAGCTC | 1080 |
| CACACCCTCT CTGACACCCT GTGGGGACCT GGTGACTCTC GGCTACAGCT GAACTTCCGA | 1140 |
| GCGACGCAGC CTTTGAATGG GCGAGTGATT GAGGCCTCCT TCCCTGCTGG AGTGGACAGC | 1200 |
| AGTCCTCGGG CTGCTGAGCC AGTCCAGCTG AATTCCTGCC TGGCTGCTGG TGACATCCTA | 1260 |
| GCCCTGGTTT TTGGCCTCCT TTTTGCTGTC ACCAGCGTCG CGTTCCTTGT GCAGATGAGA | 1320 |
| AGGCAGCACA GAAGGGGAAC CAAAGGGGGT GTGAGCTACC GCCCAGCAGA GGTAGCCGAG | 1380 |
| ACTGGAGCCT AGAGGCTGGA TCTTGAGAA TGTGAGAAGC CAGCCAGAGG CATCTGAGGG | 1440 |
| GGAGCCGGTA ACTGTCCTGT CCTGCTCATT ATGCCACTTC CTTTAACTG CCAAGAAATT | 1500 |
| TTTAAAATA AATATTTATA AT | 1522 |

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 459 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

- (A) DESCRIPTION: First 37 amino acids represent signal peptide, and remaining amino acids represent mature protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Ala | Pro | Leu | Cys | Pro | Ser | Pro | Trp | Leu | Pro | Leu | Leu | Ile | Pro | Ala |
| -35 | | | | | | | -30 | | | | | | -25 | | |

| | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|
| Pro | Ala | Pro | Gly | Leu | Thr | Val | Gln | Leu | Leu | Leu | Ser | Leu | Leu | Leu | Leu | |
| -20 | | | | | | -15 | | | | | -10 | | | | | |
| Met | Pro | Val | His | Pro | Gln | Arg | Leu | Pro | Arg | Met | Gln | Glu | Asp | Ser | Pro | |
| -5 | | | | | 1 | | | | 5 | | | | | 10 | | |
| Leu | Gly | Gly | Gly | Ser | Ser | Gly | Glu | Asp | Asp | Pro | Leu | Gly | Glu | Glu | Asp | |
| | | | 15 | | | | | 20 | | | | | 25 | | | |
| Leu | Pro | Ser | Glu | Glu | Asp | Ser | Pro | Arg | Glu | Glu | Asp | Pro | Pro | Gly | Glu | |
| | | 30 | | | | | 35 | | | | | 40 | | | | |
| Glu | Asp | Leu | Pro | Gly | Glu | Glu | Asp | Leu | Pro | Gly | Glu | Glu | Asp | Leu | Pro | |
| | 45 | | | | | 50 | | | | | 55 | | | | | |
| Glu | Val | Lys | Pro | Lys | Ser | Glu | Glu | Glu | Gly | Ser | Leu | Lys | Leu | Glu | Asp | |
| 60 | | | | | 65 | | | | | 70 | | | | | 75 | |
| Leu | Pro | Thr | Val | Glu | Ala | Pro | Gly | Asp | Pro | Gln | Glu | Pro | Gln | Asn | Asn | |
| | | | | 80 | | | | | 85 | | | | | 90 | | |
| Ala | His | Arg | Asp | Lys | Glu | Gly | Asp | Asp | Gln | Ser | His | Trp | Arg | Tyr | Gly | |
| | | | 95 | | | | | 100 | | | | | 105 | | | |
| Gly | Asp | Pro | Pro | Trp | Pro | Arg | Val | Ser | Pro | Ala | Cys | Ala | Gly | Arg | Phe | |
| | | 110 | | | | | 115 | | | | | 120 | | | | |
| Gln | Ser | Pro | Val | Asp | Ile | Arg | Pro | Gln | Leu | Ala | Ala | Phe | Cys | Pro | Ala | |
| | | 125 | | | | 130 | | | | | | 135 | | | | |
| Leu | Arg | Pro | Leu | Glu | Leu | Leu | Gly | Phe | Gln | Leu | Pro | Pro | Leu | Pro | Glu | |
| 140 | | | | | 145 | | | | | 150 | | | | | 155 | |
| Leu | Arg | Leu | Arg | Asn | Asn | Gly | His | Ser | Val | Gln | Leu | Thr | Leu | Pro | Pro | |
| | | | | 160 | | | | | 165 | | | | | 170 | | |
| Gly | Leu | Glu | Met | Ala | Leu | Gly | Pro | Gly | Arg | Glu | Tyr | Arg | Ala | Leu | Gln | |
| | | | 175 | | | | | 180 | | | | | 185 | | | |
| Leu | His | Leu | His | Trp | Gly | Ala | Ala | Gly | Arg | Pro | Gly | Ser | Glu | His | Thr | |
| | | 190 | | | | | 195 | | | | | 200 | | | | |
| Val | Glu | Gly | His | Arg | Phe | Pro | Ala | Glu | Ile | His | Val | Val | His | Leu | Ser | |
| | 205 | | | | | 210 | | | | | 215 | | | | | |

| | | | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Thr | Ala | Phe | Ala | Arg | Val | Asp | Glu | Ala | Leu | Gly | Arg | Pro | Gly | Gly | Leu | 220 | 225 | 230 | 235 |
| Ala | Val | Leu | Ala | Ala | Phe | Leu | Glu | Glu | Gly | Pro | Glu | Glu | Asn | Ser | Ala | 240 | 245 | 250 | |
| Tyr | Glu | Gln | Leu | Leu | Ser | Arg | Leu | Glu | Glu | Ile | Ala | Glu | Glu | Gly | Ser | 255 | 260 | 265 | |
| Glu | Thr | Gln | Val | Pro | Gly | Leu | Asp | Ile | Ser | Ala | Leu | Leu | Pro | Ser | Asp | 270 | 275 | 280 | |
| Phe | Ser | Arg | Tyr | Phe | Gln | Tyr | Glu | Gly | Ser | Leu | Thr | Thr | Pro | Pro | Cys | 285 | 290 | 295 | |
| Ala | Gln | Gly | Val | Ile | Trp | Thr | Val | Phe | Asn | Gln | Thr | Val | Met | Leu | Ser | 300 | 305 | 310 | 315 |
| Ala | Lys | Gln | Leu | His | Thr | Leu | Ser | Asp | Thr | Leu | Trp | Gly | Pro | Gly | Asp | 320 | 325 | 330 | |
| Ser | Arg | Leu | Gln | Leu | Asn | Phe | Arg | Ala | Thr | Gln | Pro | Leu | Asn | Gly | Arg | 335 | 340 | 345 | |
| Val | Ile | Glu | Ala | Ser | Phe | Pro | Ala | Gly | Val | Asp | Ser | Ser | Pro | Arg | Ala | 350 | 355 | 360 | |
| Ala | Glu | Pro | Val | Gln | Leu | Asn | Ser | Cys | Leu | Ala | Ala | Gly | Asp | Ile | Leu | 365 | 370 | 375 | |
| Ala | Leu | Val | Phe | Gly | Leu | Leu | Phe | Ala | Val | Thr | Ser | Val | Ala | Phe | Leu | 380 | 385 | 390 | 395 |
| Val | Gln | Met | Arg | Arg | Gln | His | Arg | Arg | Gly | Thr | Lys | Gly | Gly | Val | Ser | 400 | 405 | 410 | |
| Tyr | Arg | Pro | Ala | Glu | Val | Ala | Glu | Thr | Gly | Ala | | | | | | 415 | 420 | | |

(2) INFORMATION FOR SEQ ID NO: 3:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 29 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

CGCCCAGTGG GTCATCTTCC CCAGAAGAG

29

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GGAATCCTCC TGCATCCGG

19

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10898 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

| | |
|---|------|
| GGATCCTGTT GACTCGTGAC CTTACCCCCA ACCCTGTGCT CTCTGAAACA TGAGCTGTGT | 60 |
| CCACTCAGGG TTAAATGGAT TAAGGGCGGT GCAAGATGTG CTTTGTTAAA CAGATGCTTG | 120 |
| AAGGCAGCAT GCTCGTTAAG AGTCATCACC AATCCCTAAT CTCAAGTAAT CAGGGACACA | 180 |
| AACACTGCGG AAGGCCGCAG GGTCTCTGTC CTAGGAAAAC CAGAGACCTT TGTTCACTTG | 240 |
| TTTATCTGAC CTTCCCTCCA CTATTGTCCA TGACCCTGCC AAATCCCCCT CTGTGAGAAA | 300 |
| CACCCAAGAA TTATCAATAA AAAAATAAAT TTAATAAAAA AATACAAAAA AAAAAAAAAA | 360 |
| AAAAAAAAAA GACTTACGAA TAGTTATTGA TAAATGAATA GCTATTGGTA AAGCCAAGTA | 420 |
| AATGATCATA TTCAAACCA GACGGCCATC ATCACAGCTC AAGTCTACCT GATTTGATCT | 480 |
| CTTTATCATT GTCATTCTTT GGATTCATA GATTAGTCAT CATCCTCAA ATTCTCCCCC | 540 |
| AAGTTCTAAT TACGTTCCAA ACATTTAGGG GTTACATGAA GCTTGAACCT ACTACCTTCT | 600 |
| TTGCTTTTGA GCCATGAGTT GTAGGAATGA TGAGTTTACA CCTTACATGC TGGGGATTAA | 660 |
| TTTAACTTT ACCTCTAAGT CAGTTGGGTA GCCTTTGGCT TATTTTGTGA GCTAATTTTG | 720 |
| TAGTTAATGG ATGCACTGTG AATCTTGCTA TGATAGTTTT CCTCCACACT TTGCCACTAG | 780 |
| GGGTAGGTAG GTACTCAGTT TTCAGTAATT GCTTACCTAA GACCCTAAGC CCTATTTCTC | 840 |
| TTGTACTGGC CTTTATCTGT AATATGGGCA TATTTAATAC AATATAATTT TTGGAGTTTT | 900 |
| TTTGTTTGTT TGTTTGTTTG TTTTTTTGAG ACGGAGTCTT GCATCTGTCA TGCCCAGGCT | 960 |
| GGAGTAGCAG TGGTGCCATC TCGGCTCACT GCAAGCTCCA CCTCCCGAGT TCACGCCATT | 1020 |
| TTCTGCCTC AGCCTCCCGA GTAGCTGGGA CTACAGGCGC CGGCCACCAT GCCCGGCTAA | 1080 |

| | | | | | | |
|------------|------------|------------|------------|-------------|------------|------|
| TTTTTTGTAT | TTTTGGTAGA | GACGGGGTTT | CACCGTGTTA | GCCAGAAATGG | TCTCGATCTC | 1140 |
| CTGACTTCGT | GATCCACCCG | CCTCGGCCTC | CCAAAGTTCT | GGGATTACAG | GTGTGAGCCA | 1200 |
| CCGCACCTGG | CCAATTTTTT | GAGTCTTTTA | AAGTAAAAAT | ATGTCTTGTA | AGCTGGTAAC | 1260 |
| TATGGTACAT | TTCCTTTTAT | TAATGTGGTG | CTGACGGTCA | TATAGGTTCT | TTTGAGTTTG | 1320 |
| GCATGCATAT | GCTACTTTTT | GCAGTCCTTT | CATTACATTT | TTCTCTCTTC | ATTTGAAGAG | 1380 |
| CATGTTATAT | CTTTTAGCTT | CACCTGGCTT | AAAAGGTTCT | CTCATTAGCC | TAACACAGTG | 1440 |
| TCATTGTTGG | TACCACTTGG | ATCATAAGTG | GAAAAACAGT | CAAGAAATTG | CACAGTAATA | 1500 |
| CTTGTTTGTA | AGAGGGATGA | TTCAGGTGAA | TCTGACACTA | AGAAACTCCC | CTACCTGAGG | 1560 |
| TCTGAGATTC | CTCTGACATT | GCTGTATATA | GGCTTTTCCT | TTGACAGCCT | GTGACTGCGG | 1620 |
| ACTATTTTTT | TTAAGCAAGA | TATGCTAAAG | TTTTGTGAGC | CTTTTTCCAG | AGAGAGGTCT | 1680 |
| CATATCTGCA | TCAAGTGAGA | ACATATAATG | TCTGCATGTT | TCCATATTTC | AGGAATGTTT | 1740 |
| GCTTGTGTTT | TATGCTTTTA | TATAGACAGG | GAAACTTGTT | CCTCAGTGAC | CCAAAAGAGG | 1800 |
| TGGGAATTGT | TATTGGATAT | CATCATTGGC | CCACGCTTTC | TGACCTTGGA | AACAATTAAG | 1860 |
| GGTTCATAAT | CTCAATTCTG | TCAGAATTGG | TACAAGAAAT | AGCTGCTATG | TTTCTTGACA | 1920 |
| TTCCACTTGG | TAGGAAATAA | GAATGTGAAA | CTCTTCAGTT | GGTGTGTGTC | CCTNGTTTTT | 1980 |
| TTGCAATTTT | CTTCTTACTG | TGTTAAAAAA | AAGTATGATC | TTGCTCTGAG | AGGTGAGGCA | 2040 |
| TTCTTAATCA | TGATCTTTAA | AGATCAATAA | TATAATCCTT | TCAAGGATTA | TGTCTTTATT | 2100 |
| ATAATAAAGA | TAATTTGTCT | TTAACAGAAT | CAATAATATA | ATCCCTTAAA | GGATTATATC | 2160 |
| TTTGCTGGGC | GCAGTGGCTC | ACACCTGTAA | TCCCAGCACT | TTGGGTGGCC | AAGGTGGAAG | 2220 |
| GATCAAATTT | GCCTACTTCT | ATATTATCTT | CTAAAGCAGA | ATTCATCTCT | CTTCCCTCAA | 2280 |
| TATGATGATA | TTGACAGGGT | TTGCCCTCAC | TCTAGATT | GTGAGCTCCT | GCTCAGGGCA | 2340 |
| GGTAGCGTTT | TTTGTTTTTG | TTTTTGTTTT | TCTTTTTTGA | GACAGGGTCT | TGCTCTGTCA | 2400 |
| CCCAGGCCAG | AGTGCAATGG | TACAGTCTCA | GCTCACTGCA | GCCTCAACCG | CCTCGGCTCA | 2460 |

| | | | | | | |
|------------|------------|------------|------------|------------|-------------|------|
| AACCATCATC | CCATTTCAGC | CTCCTGAGTA | GCTGGGACTA | CAGGCACATG | CCATTACACC | 2520 |
| TGGCTAATTT | TTTTGTATTT | CTAGTAGAGA | CAGGGTTTGG | CCATGTTGCC | CGGGCTGGTC | 2580 |
| TCGAACTCCT | GGACTCAAGC | AATCCACCCA | CCTCAGCCTC | CCAAAATGAG | GGACCGTGTC | 2640 |
| TTATTCATTT | CCATGTCCCT | AGTCCATAGC | CCAGTGCTGG | ACCTATGGTA | GTACTAAATA | 2700 |
| AATATTTGTT | GAATGCAATA | GTAAATAGCA | TTTCAGGGAG | CAAGAACTAG | ATTAACAAAG | 2760 |
| GTGGTAAAAG | GTTTGGAGAA | AAAAATAATA | GTTTAATTTG | GCTAGAGTAT | GAGGGAGAGT | 2820 |
| AGTAGGAGAC | AAGATGGAAA | GGTCTCTTGG | GCAAGGTTTT | GAAGGAAGTT | GGAAGTCAGA | 2880 |
| AGTACACAAT | GTGCATATCG | TGGCAGGCAG | TGGGGAGCCA | ATGAAGGCTT | TTGAGCAGGA | 2940 |
| GAGTAATGTG | TTGAAAAATA | AATATAGGTT | AAACCTATCA | GAGCCCCTCT | GACACATACA | 3000 |
| CTTGCTTTTC | ATTCAAGCTC | AAGTTTGTCT | CCCACATACC | CATTACTTAA | CTCACCCCTCG | 3060 |
| GGCTCCCCTA | GCAGCCTGCC | CTACCTCTTT | ACCTGCTTCC | TGGTGGAGTC | AGGGATGTAT | 3120 |
| ACATGAGCTG | CTTTCCTCT | CAGCCAGAGG | ACATGGGGGG | CCCCAGCTCC | CCTGCCTTTC | 3180 |
| CCCTTCTGTG | CCTGGAGCTG | GGAAGCAGGC | CAGGGTTAGC | TGAGGCTGGC | TGGCAAGCAG | 3240 |
| CTGGGTGGTG | CCAGGGAGAG | CCTGCATAGT | GCCAGGTGGT | GCCTTGGGTT | CCAAGCTAGT | 3300 |
| CCATGGCCCC | GATAACCTTC | TGCCTGTGCA | CACACCTGCC | CCTCACTCCA | CCCCCATCCT | 3360 |
| AGCTTTGGTA | TGGGGGAGAG | GGCACAGGGC | CAGACAAACC | TGTGAGACTT | TGGCTCCATC | 3420 |
| TCTGCAAAAG | GGCGCTCTGT | GAGTCAGCCT | GCTCCCCTCC | AGGCTTGCTC | CTCCCCCACC | 3480 |
| CAGCTCTCGT | TTCCAATGCA | CGTACAGCCC | GTACACACCG | TGTGCTGGGA | CACCCCACAG | 3540 |
| TCAGCCGCAT | GGCTCCCCTG | TGCCCCAGCC | CCTGGCTCCC | TCTGTTGATC | CCGGCCCCCTG | 3600 |
| CTCCAGGCCT | CACTGTGCAA | CTGCTGCTGT | CACTGCTGCT | TCTGGTGCCT | GTCCATCCCC | 3660 |
| AGAGGTTGCC | CCGGATGCAG | GAGGATTCCC | CCTTGGGAGG | AGGCTCTTCT | GGGGAAGATG | 3720 |
| ACCCACTGGG | CGAGGAGGAT | CTGCCCAGTG | AAGAGGATTC | ACCCAGAGAG | GAGGATCCAC | 3780 |
| CCGGAGAGGA | GGATCTACCT | GGAGAGGAGG | ATCTACCTGG | AGAGGAGGAT | CTACCTGAAG | 3840 |

| | | | | | | |
|------------|------------|------------|------------|------------|------------|------|
| TTAAGCCTAA | ATCAGAAGAA | GAGGGCTCCC | TGAAGTTAGA | GGATCTACCT | ACTGTTGAGG | 3900 |
| CTCCTGGAGA | TCCTCAAGAA | CCCCAGAATA | ATGCCCACAG | GGACAAAGAA | GGTAAGTGGT | 3960 |
| CATCAATCTC | CAAATCCAGG | TTCCAGGAGG | TTCATGACTC | CCCTCCCATA | CCCCAGCCTA | 4020 |
| GGCTCTGTTT | ACTCAGGGAA | GGAGGGGAGA | CTGTACTCCC | CACAGAAGCC | CTTCCAGAGG | 4080 |
| TCCCATACCA | ATATCCCCAT | CCCCACTCTC | GGAGGTAGAA | AGGGACAGAT | GTGGAGAGAA | 4140 |
| AATAAAAAGG | GTGCAAAAGG | AGAGAGGTGA | GCTGGATGAG | ATGGGAGAGA | AGGGGGAGGC | 4200 |
| TGGAGAAGAG | AAAGGGATGA | GAAGTGCAGA | TGAGAGAAAA | AATGTGCAGA | CAGAGGAAAA | 4260 |
| AAATAGGTGG | AGAAGGAGAG | TCAGAGAGTT | TGAGGGGAAG | AGAAAAGGAA | AGCTTGGGAG | 4320 |
| GTGAAGTGGG | TACCAGAGAC | AAGCAAGAA | AGCTGGTAGA | AGTCATCTCA | TCTTAGGCTA | 4380 |
| CAATGAGGAA | TTGAGACCTA | GGAAGAAGGG | ACACAGCAGG | TAGAGAAACG | TGGCTTCTTG | 4440 |
| ACTCCCAAGC | CAGGAATTTG | GGGAAAGGGG | TTGGAGACCA | TACAAGGCAG | AGGGATGAGT | 4500 |
| GGGGAGAAGA | AAGAAGGGAG | AAAGGAAAGA | TGGTGTACTC | ACTCATTTGG | GAATCAGGAC | 4560 |
| TGAAGTGCCC | ACTCACTTTT | TTTTTTTTTT | TTTTTGAGAC | AAACTTTCAC | TTTTGTTGCC | 4620 |
| CAGGCTGGAG | TGCAATGGCG | CGATCTCGGC | TCACTGCAAC | CTCCACCTCC | CGGGTTCAAG | 4680 |
| TGATTCTCCT | GCCTCAGCCT | CTAGCCAAGT | AGCTGCGATT | ACAGGCATGC | GCCACCACGC | 4740 |
| CCGGCTAATT | TTTGTATTTT | TAGTAGAGAC | GGGGTTTCGC | CATGTTGGTC | AGGCTGGTCT | 4800 |
| CGAACTCCTG | ATCTCAGGTG | ATCCAACCAC | CCTGGCCTCC | CAAAGTGCTG | GGATTATAGG | 4860 |
| CGTGAGCCAC | AGCGCCTGGC | CTGAAGCAGC | CACTCACTTT | TACAGACCCT | AAGACAATGA | 4920 |
| TTGCAAGCTG | GTAGGATTGC | TGTTTGGCCC | ACCCAGCTGC | GGTGTGAGT | TTGGGTGCGG | 4980 |
| TCTCCTGTGC | TTTGCACCTG | GCCCCGCTAA | GGCATTGTGT | ACCCGTAATG | CTCCTGTAAG | 5040 |
| GCATCTGCGT | TTGTGACATC | GTTTTGGTGC | CCAGGAAGGG | ATTGGGGCTC | TAAGCTTGAG | 5100 |
| CGGTTCATCC | TTTTCATTTA | TACAGGGGAT | GACCAGAGTC | ATTGGCGCTA | TGGAGGTGAG | 5160 |
| ACACCCACCC | GCTGCACAGA | CCCAATCTGG | GAACCCAGCT | CTGTGGATCT | CCCCTACAGC | 5220 |

| | | | | | | |
|------------|------------|------------|------------|------------|------------|------|
| CGTCCCTGAA | CACTGGTCCC | GGGCGTCCCA | CCCGCCGCCC | ACCGTCCCAC | CCCCTCACCT | 5280 |
| TTTCTACCCG | GGTTCCTAA | GTTCTGACC | TAGGCGTCAG | ACTTCCTCAC | TATACTCTCC | 5340 |
| CACCCAGGC | GACCCGCCCT | GGCCCCGGGT | GTCCCCAGCC | TGCGCGGGCC | GCTTCCAGTC | 5400 |
| CCCGGTGGAT | ATCCGCCCCC | AGCTCGCCGC | CTTCTGCCCC | GCCCTGCGCC | CCCTGGAACT | 5460 |
| CCTGGGCTTC | CAGCTCCCGC | CGCTCCCAGA | ACTGCGCCTG | CGCAACAATG | GCCACAGTGG | 5520 |
| TGAGGGGGTC | TCCCCGCCGA | GACTTGGGGA | TGGGGCGGGG | CGCAGGGAAG | GGAACCGTCG | 5580 |
| CGCAGTGCCT | GCCCCGGGGT | TGGGCTGGCC | CTACCGGGCG | GGGCCGGCTC | ACTTGCCCTT | 5640 |
| CCCTACGCAG | TGCAACTGAC | CCTGCCTCCT | GGGCTAGAGA | TGGCTCTGGG | TCCCGGGCGG | 5700 |
| GAGTACCGGG | CTCTGCAGCT | GCATCTGCAC | TGGGGGGCTG | CAGGTCGTCC | GGGCTCGGAG | 5760 |
| CACACTGTGG | AAGGCCACCG | TTTCCCTGCC | GAGGTGAGCG | CGGACTGGCC | GAGAAGGGGC | 5820 |
| AAAGGAGCGG | GGCGGACGGG | GGCCAGAGAC | GTGGCCCTCT | CCTACCCTCG | TGTCCTTTTC | 5880 |
| AGATCCACGT | GGTTCACCTC | AGCACCGCCT | TTGCCAGAGT | TGACGAGGCC | TTGGGGCGCC | 5940 |
| CGGGAGGCCT | GGCCGTGTTG | GCCGCCTTTC | TGGAGGTACC | AGATCCTGGA | CACCCCCTAC | 6000 |
| TCCCCGCTTT | CCCATCCCAT | GCTCCTCCCG | GACTCTATCG | TGGAGCCAGA | GACCCCATCC | 6060 |
| CAGCAAGCTC | ACTCAGGCCC | CTGGCTGACA | AACTCATTCA | CGCACTGTTT | GTTCAATTAA | 6120 |
| CACCCACTGT | GAACCAGGCA | CCAGCCCCCA | ACAAGGATTG | TGAAGCTGTA | GGTCCTTGCC | 6180 |
| TCTAAGGAGC | CCACAGCCAG | TGGGGGAGGC | TGACATGACA | GACACATAGG | AAGGACATAG | 6240 |
| TAAAGATGGT | GGTCACAGAG | GAGGTGACAC | TTAAAGCCTT | CACTGGTAGA | AAAGAAAAGG | 6300 |
| AGGTGTTTCA | TGCAGAGGAA | ACAGAATGTG | CAAAGACTCA | GAATATGGCC | TATTTAGGGA | 6360 |
| ATGGCTACAT | ACACCATGAT | TAGAGGAGGC | CCAGTAAAGG | GAAGGGATGG | TGAGATGCCT | 6420 |
| GCTAGGTTCA | CTCACTCACT | TTTATTTATT | TATTTATTTT | TTTGACAGTC | TCTCTGTCGC | 6480 |
| CCAGGCTGGA | GTGCAGTGGT | GTGATCTTGG | GTCAGTGCAA | CTTCGGCCTC | CCGGGTTCAG | 6540 |
| GGGATTCTCC | TGCCTCAGCT | TCCTGAGTAG | CTGGGGTTAC | AGGTGTGTGC | CACCATGCCC | 6600 |

| | | | | | | |
|-------------|------------|-------------|------------|------------|------------|------|
| AGCTAATTTT | TTTTTGTATT | TTTAGTAGAC | AGGGTTTCAC | CATGTTGGTC | AGGCTGGTCT | 6660 |
| CAAACCTCCTG | GCCTCAAGTG | ATCCGCCTGA | CTCAGCCTAC | CAAAGTGCTG | ATTACAAGTG | 6720 |
| TGAGCCACCG | TGCCCAGCCA | CACTCACTGA | TTCTTTAATG | CCAGCCACAC | AGCACAAAGT | 6780 |
| TCAGAGAAAT | GCCTCCATCA | TAGCATGTCA | ATATGTTTAT | ACTCTTAGGT | TCATGATGTT | 6840 |
| CTTAACATTA | GGTTCATAAG | CAAAATAAGA | AAAAAGAATA | ATAAATAAAA | GAAGTGGCAT | 6900 |
| GTCAGGACCT | CACCTGAAAA | GCCAAACACA | GAATCATGAA | GGTGAATGCA | GAGGTGACAC | 6960 |
| CAACACAAAG | GTGTATATAT | GGTTTCCTGT | GGGGAGTATG | TACGGAGGCA | GCAGTGAGTG | 7020 |
| AGACTGCAAA | CGTCAGAAGG | GCACGGGTCA | CTGAGAGCCT | AGTATCCTAG | TAAAGTGGGC | 7080 |
| TCTCTCCCTC | TCTCTCCAGC | TTGTCAATTGA | AAACCAGTCC | ACCAAGCTTG | TTGGTTCGCA | 7140 |
| CAGCAAGAGT | ACATAGAGTT | TGAAATAATA | CATAGGATTT | TAAGAGGGAG | ACACTGTCTC | 7200 |
| TAAAAAATA | AACAACAGCA | ACAACAAAAA | GCAACAACCA | TTACAATTTT | ATGTTCCCTC | 7260 |
| AGCATTCTCA | GAGCTGAGGA | ATGGGAGAGG | ACTATGGGAA | CCCCCTTCAT | GTTCCGGCCT | 7320 |
| TCAGCCATGG | CCCTGGATAC | ATGCACTCAT | CTGTCTTACA | ATGTCATTCC | CCCAGGAGGG | 7380 |
| CCCGGAAGAA | AACAGTGCCT | ATGAGCAGTT | GCTGTCTCGC | TTGGAAGAAA | TCGCTGAGGA | 7440 |
| AGGTCAGTTT | GTTGGTCTGG | CCACTAATCT | CTGTGGCCTA | GTCATAAAG | AATCACCTTT | 7500 |
| TGGAGCTTCA | GGTCTGAGGC | TGGAGATGGG | CTCCTCCAG | TGCAGGAGGG | ATTGAAGCAT | 7560 |
| GAGCCAGCGC | TCATCTTGAT | AATAACCATG | AAGCTGACAG | ACACAGTTAC | CCGCAAACGG | 7620 |
| CTGCCTACAG | ATTGAAAACC | AAGCAAAAAC | CGCCGGGCAC | GGTGGCTCAC | GCCTGTAATC | 7680 |
| CCAGCACTTT | GGGAGGCCAA | GGCAGGTGGA | TCACGAGGTC | AAGAGATCAA | GACCATCCTG | 7740 |
| GCCAACATGG | TGAAACCCCA | TCTCTACTAA | AAATACGAAA | AAATAGCCAG | GCGTGGTGGC | 7800 |
| GGGTGCCTGT | AATCCCAGCT | ACTCGGGAGG | CTGAGGCAGG | AGAATGGCAT | GAACCCGGGA | 7860 |
| GGCAGAAGTT | GCAGTGAGCC | GAGATCGTGC | CACTGCACTC | CAGCCTGGGC | AACAGAGCGA | 7920 |
| GACTCTTGTC | TCAAAAAAAA | AAAAAATAAA | GAAAACCAAG | CAAAAACCAA | AATGAGACAA | 7980 |

| | | | | | | |
|------------|-------------|------------|------------|------------|------------|------|
| AAAAAACAAG | ACCAAAAAAT | GGTGTTTGGA | AATTGTCAAG | GTCAAGTCTG | GAGAGCTAAA | 8040 |
| CTTTTTCTGA | GAAGTGTTTA | TCTTTAATAA | GCATCAAATA | TTTAACTTT | GTAAATACTT | 8100 |
| TTGTTGGAAA | TCGTTCTCTT | CTTAGTCACT | CTTGGGTCAT | TTTAAATCTC | ACTTACTCTA | 8160 |
| CTAGACCTTT | TAGGTTTCTG | CTAGACTAGG | TAGAACTCTG | CCTTTGCATT | TCTTGTGTCT | 8220 |
| GTTTTGTATA | GTTATCAATA | TTCATATTTA | TTTACAAGTT | ATTCAGATCA | TTTTTTCTTT | 8280 |
| TCTTTTTTTT | TTTTTTTTTT | TTTTTTACAT | CTTTAGTAGA | GACAGGGTTT | CACCATATTG | 8340 |
| GCCAGGCTGC | TCTCAAACCTC | CTGACCTTGT | GATCCACCAG | CCTCGGCCTC | CCAAAGTGCT | 8400 |
| GGGATTCATT | TTTTCTTTTT | AATTTGCTCT | GGGCTTAAAC | TTGTGGCCCA | GCACTTTATG | 8460 |
| ATGGTACACA | GAGTTAAGAG | TGTAGACTCA | GACGGTCTTT | CTTCTTTCCT | TCTCTTCCTT | 8520 |
| CCTCCCTTCC | CTCCCACCTT | CCCTTCTCTC | CTTCCTTTCT | TTCTTCCTCT | CTTGCTTCCT | 8580 |
| CAGGCCTCTT | CCAGTTGCTC | CAAAGCCCTG | TACTTTTTTT | TGAGTTAACG | TCTTATGGGA | 8640 |
| AGGCCTTGCA | CTTAGTGAAG | AAGTGGTCTC | AGAGTTGAGT | TACCTTGGCT | TCTGGGAGGT | 8700 |
| GAAACTGTAT | CCCTATACCC | TGAAGCTTTA | AGGGGGTGCA | ATGTAGATGA | GACCCCAACA | 8760 |
| TAGATCCTCT | TCACAGGCTC | AGAGACTCAG | GTCCGAGGAC | TGGACATATC | TGCACTCCTG | 8820 |
| CCCTCTGACT | TCAGCCGCTA | CTTCCAATAT | GAGGGGTCTC | TGACTACACC | GCCCTGTGCC | 8880 |
| CAGGGTGTCA | TCTGGACTGT | GTTTAACCAG | ACAGTGATGC | TGAGTGCTAA | GCAGGTGGGC | 8940 |
| CTGGGGTGTG | TGTGGACACA | GTGGGTGCGG | GGGAAAGAGG | ATGTAAGATG | AGATGAGAAA | 9000 |
| CAGGAGAAGA | AAGAAATCAA | GGCTGGGCTC | TGTGGCTTAC | GCCTATAATC | CCACCACGTT | 9060 |
| GGGAGGCTGA | GGTGGGAGAA | TGGTTTGAGC | CCAGGAGTTC | AAGACAAGGC | GGGGCAACAT | 9120 |
| AGTGTGACCC | CATCTCTACC | AAAAAAACCC | CAACAAAACC | AAAAATAGCC | GGGCATGGTG | 9180 |
| GTATGCGGCC | TAGTCCCAGC | TACTCAAGGA | GGCTGAGGTG | GGAAGATCGC | TTGATTCCAG | 9240 |
| GAGTTTGAGA | CTGCAGTGAG | CTATGATCCC | ACCACTGCCT | ACCATCTTTA | GGATACATTT | 9300 |
| ATTTATTTAT | AAAAGAAATC | AAGAGGCTGG | ATGGGGAATA | CAGGAGCTGG | AGGGTGGAGC | 9360 |

| | | | | | | |
|------------|-------------|------------|------------|------------|------------|-------|
| CCTGAGGTGC | TGGTTGTGAG | CTGGCCTGGG | ACCCTTGTTT | CCTGTCATGC | CATGAACCCA | 9420 |
| CCCACACTGT | CCACTGACCT | CCCTAGCTCC | ACACCCTCTC | TGACACCCTG | TGGGGACCTG | 9480 |
| GTGACTCTCG | GCTACAGCTG | AACTTCCGAG | CGACGCAGCC | TTTGAATGGG | CGAGTGATTG | 9540 |
| AGGCCTCCTT | CCCTGCTGGA | GTGGACAGCA | GTCCTCGGGC | TGCTGAGCCA | GGTACAGCTT | 9600 |
| TGTCTGGTTT | CCCCCAGCC | AGTAGTCCCT | TATCCTCCCA | TGTGTGTGCC | AGTGTCTGTC | 9660 |
| ATTGGTGGTC | ACAGCCCGCC | TCTCACATCT | CCTTTTTCTC | TCCAGTCCAG | CTGAATTCCT | 9720 |
| GCCTGGCTGC | TGGTGAGTCT | GCCCCTCCTC | TTGGTCCTGA | TGCCAGGAGA | CTCCTCAGCA | 9780 |
| CCATTCAGCC | CCAGGGCTGC | TCAGGACCGC | CTCTGCTCCC | TCTCCTTTTC | TGCAGAACAG | 9840 |
| ACCCCAACCC | CAATATTAGA | GAGGCAGATC | ATGGTGGGGA | TTCCCCCATT | GTCCCCAGAG | 9900 |
| GCTAATTGAT | TAGAATGAAG | CTTGAGAAAT | CTCCCAGCAT | CCCTCTCGCA | AAAGAATCCC | 9960 |
| CCCCCCTTTT | TTTAAAGATA | GGGTCTCACT | CTGTTTGCCC | CAGGCTGGGG | TGTTGTGGCA | 10020 |
| CGATCATAGC | TCACTGCAGC | CTCGAACTCC | TAGGCTCAGG | CAATCCTTTC | ACCTTAGCTT | 10080 |
| CTCAAAGCAC | TGGGACTGTA | GGCATGAGCC | ACTGTCCCTG | GCCCCAAACG | GCCCTTTTAC | 10140 |
| TTGGCTTTTA | GGAAGCAAAA | ACGGTGCTTA | TCTTACCCCT | TCTCGTGTAT | CCACCCTCAT | 10200 |
| CCCTTGGCTG | GCCTCTTCTG | GAGACTGAGG | CACTATGGGG | CTGCCTGAGA | ACTCGGGGCA | 10260 |
| GGGGTGGTGG | AGTGCACTGA | GGCAGGTGTT | GAGGAACTCT | GCAGACCCCT | CTTCCTTCCC | 10320 |
| AAAGCAGCCC | TCTCTGCTCT | CCATCGCAGG | TGACATCCTA | GCCCTGGTTT | TTGGCCTCCT | 10380 |
| TTTTGCTGTC | ACCAGCGTCG | CGTTCCTTGT | GCAGATGAGA | AGGCAGCACA | GGTATTACAC | 10440 |
| TGACCCTTTC | TTCAGGCACA | AGCTTCCCCC | ACCCTTGTGG | AGTCACTTCA | TGCAAAGCGC | 10500 |
| ATGCAAATGA | GCTGCTCCTG | GGCCAGTTTT | CTGATTAGCC | TTTCCTGTTG | TGTACACACA | 10560 |
| GAAGGGGAAC | CAAAGGGGGT | GTGAGCTACC | GCCCAGCAGA | GGTAGCCGAG | ACTGGAGCCT | 10620 |
| AGAGGCTGGA | TCTTGAGAGAA | TGTGAGAAGC | CAGCCAGAGG | CATCTGAGGG | GGAGCCGGTA | 10680 |
| ACTGTCCTGT | CCTGCTCATT | ATGCCACTTC | CTTTTAACTG | CCAAGAAATT | TTTTAAAATA | 10740 |

AATATTTATA ATAAAATATG TGTTAGTCAC CTTTGTTCCC CAAATCAGAA GGAGGTATTT 10800
 GAATTTCTTA TTACTGTTAT TAGCACCAAT TTAGTGGTAA TGCATTTATT CTATTACAGT 10860
 TCGGCCTCCT TCCACACATC ACTCCAATGT GTTGCTCC 10898

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (A) DESCRIPTION: Signal peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Ala | Pro | Leu | Cys | Pro | Ser | Pro | Trp | Leu | Pro | Leu | Leu | Ile | Pro | Ala |
| 1 | | | | 5 | | | | | 10 | | | | | 15 | |
| Pro | Ala | Pro | Gly | Leu | Thr | Val | Gln | Leu | Leu | Leu | Ser | Leu | Leu | Leu | Leu |
| | | | 20 | | | | 25 | | | | | 30 | | | |
| Met | Pro | Val | His | Pro | | | | | | | | | | | |
| | | | 35 | | | | | | | | | | | | |

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

TGGGGTTCTT GAGGATCTCC AGGAG

25

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "primer"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

CTCTAACTTC AGGGAGCCCT CTTCTT

26

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "primer"

(iii) HYPOTHETICAL: NO

(ix) FEATURE:

(D) OTHER INFORMATION: N stands for inosine

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

CUACUACUAC UAGGCCACGC GTCGACTAGT ACGGGNNGGG NNGGGNNG

48

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Glu Glu Asp Leu Pro Ser

1

5

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(ix) FEATURE:

(A) NAME/KEY: Peptide

(B) LOCATION:55..60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

Gly Glu Asp Asp Pro Leu
1 5

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Asn Asn Ala His Arg Asp Lys Glu Gly Asp Asp Gln Ser His Trp Arg
1 5 10 15

Tyr Gly Gly Asp Pro
20

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(ix) FEATURE:

- (A) NAME/KEY: Peptide
- (B) LOCATION: 36..51

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

His Pro Gln Arg Leu Pro Arg Met Gln Glu Asp Ser Pro Leu Gly Gly
1 5 10 15

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Glu Glu Asp Ser Pro Arg Glu Glu Asp Pro Pro Gly Glu Glu Asp Leu
1 5 10 15

Pro Gly Glu Glu Asp Leu Pro Gly
20

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(ix) FEATURE:

- (A) NAME/KEY: Peptide
- (B) LOCATION: 279..291

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

Leu Glu Glu Gly Pro Glu Glu Asn Ser Ala Tyr Glu Gln
1 5 10

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Met Arg Arg Gln His Arg Arg Gly Thr Lys Gly Gly Val Ser Tyr Arg
1 5 10 15

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

GTCGCTAGCT CCATGGGTCA TATGCAGAGG TTGCCCCGGA TGCAG

45

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 43 base pairs
- (B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

GAAGATCTCT TACTCGAGCA TTCTCCAAGA TCCAGCCTCT AGG

43

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: DNA (genomic)

(A) DESCRIPTION: AP-2 transcription factor

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TCCCCCACCC

10

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: DNA (genomic)

(A) DESCRIPTION: initiator (Inr) element

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CCACCCCCAT

10

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: DNA (genomic)

- (A) DESCRIPTION: p53 binding site

(x) PUBLICATION INFORMATION:

- (A) AUTHORS: El Deiry et al.
- (B) TITLE: "Human genomic DNA sequences define a consensus binding site for p53"
- (C) JOURNAL: Nature Genetics
- (D) VOLUME: 1
- (F) PAGES: 44-49
- (G) DATE: 1992

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

AAGCTAGTCC

10

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

Leu Glu His His His His His His
1 5

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: Initiator consensus sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

YYYCAYYYYY

10

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: p53 binding site

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(x) PUBLICATION INFORMATION:

- (A) AUTHORS: El Deiry et al.
- (B) TITLE: "Human genomic DNA sequences define a consensus binding site for p53"
- (C) JOURNAL: Nature Genetics
- (D) VOLUME: 1
- (F) PAGES: 44-49
- (G) DATE: 1992

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

AGGCTTGCTC

10

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

Ser Pro Xaa Xaa
1

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

Thr Pro Xaa Xaa
1

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 540 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: Proposed MN promoter

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

| | |
|---|-----|
| CTTGCTTTTC ATTCAAGCTC AAGTTTGTCT CCCACATACC CATTACTTAA CTCACCCTCG | 60 |
| GGCTCCCCTA GCAGCCTGCC CTACCTCTTT ACCTGCTTCC TGGTGGAGTC AGGGATGTAT | 120 |
| ACATGAGCTG CTTTCCCTCT CAGCCAGAGG ACATGGGGGG CCCAGCTCC CCTGCCTTTC | 180 |
| CCCTTCTGTG CCTGGAGCTG GGAAGCAGGC CAGGGTTAGC TGAGGCTGGC TGGCAAGCAG | 240 |
| CTGGGTGGTG CCAGGGAGAG CCTGCATAGT GCCAGGTGGT GCCTTGGGTT CCAAGCTAGT | 300 |
| CCATGGCCCC GATAACCTTC TGCCTGTGCA CACACCTGCC CCTCACTCCA CCCCCATCCT | 360 |
| AGCTTTGGTA TGGGGGAGAG GGCACAGGGC CAGACAAACC TGTGAGACTT TGGCTCCATC | 420 |
| TCTGCAAAAG GGCGCTCTGT GAGTCAGCCT GCTCCCCTCC AGGCTTGCTC CTCCCCCACC | 480 |
| CAGCTCTCGT TTCCAATGCA CGTACAGCCC GTACACACCG TGTGCTGGGA CACCCCACAG | 540 |

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 415 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 1st MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

| | |
|---|-----|
| ACAGTCAGCC GCATGGCTCC CCTGTGCCCC AGCCCCTGGC TCCCTCTGTT GATCCCGGCC | 60 |
| CCTGCTCCAG GCCTCACTGT GCAACTGCTG CTGTCACTGC TGCTTCTGGT GCCTGTCCAT | 120 |
| CCCCAGAGGT TGCCCCGGAT GCAGGAGGAT TCCCCCTTGG GAGGAGGCTC TTCTGGGGAA | 180 |
| GATGACCCAC TGGGCGAGGA GGATCTGCCC AGTGAAGAGG ATTCACCCAG AGAGGAGGAT | 240 |
| CCACCCGGAG AGGAGGATCT ACCTGGAGAG GAGGATCTAC CTGGAGAGGA GGATCTACCT | 300 |
| GAAGTTAAGC CTAAATCAGA AGAAGAGGGC TCCCTGAAGT TAGAGGATCT ACCTACTGTT | 360 |
| GAGGCTCCTG GAGATCCTCA AGAACCCGAG AATAATGCCC ACAGGGACAA AGAAG | 415 |

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 2nd MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

| | |
|----------------------------------|----|
| GGGATGACCA GAGTCATTGG CGCTATGGAG | 30 |
|----------------------------------|----|

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 171 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 3rd MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

| | |
|---|-----|
| GCGACCCGCC CTGGCCCCGG GTGTCCCCAG CCTGCGCGGG CCGCTTCCAG TCCCCGGTGG | 60 |
| ATATCCGCCC CCAGCTCGCC GCCTTCTGCC CGGCCCTGCG CCCCTGGAA CTCCTGGGCT | 120 |
| TCCAGCTCCC GCCGCTCCCA GAACTGCGCC TGC GCAACAA TGGCCACAGT G | 171 |

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 143 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 4th MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

| | |
|---|-----|
| TGCAACTGAC CCTGCCTCCT GGGCTAGAGA TGGCTCTGGG TCCCGGGCGG GAGTACCGGG | 60 |
| CTCTGCAGCT GCATCTGCAC TGGGGGGGCTG CAGGTCGTCC GGGCTCGGAG CAACTGTGG | 120 |

AAGGCCACCG TTTCCCTGCC GAG

143

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 93 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 5th MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

ATCCACGTGG TTCACCTCAG CACCGCCTTT GCCAGAGTTG ACGAGGCCTT GGGGCGCCCG 60

GGAGGCCTGG CCGTGTTGGC CGCCTTTCTG GAG 93

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 67 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 6th MN exon

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

GAGGGCCCGG AAGAAAACAG TGCCTATGAG CAGTTGCTGT CTCGCTTGGA AGAAATCGCT 60

GAGGAAG 67

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 158 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 7th MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

| | |
|---|-----|
| GCTCAGAGAC TCAGGTCCCA GGACTGGACA TATCTGCACT CCTGCCCTCT GACTTCAGCC | 60 |
| GCTACTTCCA ATATGAGGGG TCTCTGACTA CACCGCCCTG TGCCCAGGGT GTCATCTGGA | 120 |
| CTGTGTTTAA CCAGACAGTG ATGCTGAGTG CTAAGCAG | 158 |

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 145 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 8th MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

CTCCACACCC TCTCTGACAC CCTGTGGGGA CCTGGTGACT CTCGGCTACA GCTGAACTTC 60
CGAGCGACGC AGCCTTTGAA TGGGCGAGTG ATTGAGGCCT CCTTCCCTGC TGGAGTGGAC 120
AGCAGTCCTC GGGCTGCTGA GCCAG 145

(2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 9th MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

TCCAGCTGAA TTCCTGCCTG GCTGCTG 27

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 82 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 10th MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

| | |
|---|----|
| GTGACATCCT AGCCCTGGTT TTTGGCCTCC TTTTGTCTGT CACCAGCGTC GCGTTCCTTG | 60 |
| TGCAGATGAG AAGGCAGCAC AG | 82 |

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 191 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 11th MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

| | |
|---|-----|
| AAGGGGAACC AAAGGGGGTG TGAGCTACCG CCCAGCAGAG GTAGCCGAGA CTGGAGCCTA | 60 |
| GAGGCTGGAT CTTGGAGAAT GTGAGAAGCC AGCCAGAGGC ATCTGAGGGG GAGCCGGTAA | 120 |
| CTGTCCTGTC CTGCTCATTA TGCCACTTCC TTTTAACTGC CAAGAAATTT TTTAAAATAA | 180 |
| ATATTTATAA T | 191 |

(2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1174 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 1st MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

| | |
|---|------|
| GTAAGTGGTC ATCAATCTCC AAATCCAGGT TCCAGGAGGT TCATGACTCC CCTCCCATAC | 60 |
| CCCAGCCTAG GCTCTGTTCA CTCAGGGAAG GAGGGGAGAC TGTACTCCCC ACAGAAGCCC | 120 |
| TTCCAGAGGT CCCATACCAA TATCCCCATC CCCACTCTCG GAGGTAGAAA GGGACAGATG | 180 |
| TGGAGAGAAA ATAAAAAGGG TGCAAAAGGA GAGAGGTGAG CTGGATGAGA TGGGAGAGAA | 240 |
| GGGGGAGGCT GGAGAAGAGA AAGGGATGAG AACTGCAGAT GAGAGAAAAA ATGTGCAGAC | 300 |
| AGAGGAAAAA AATAGGTGGA GAAGGAGAGT CAGAGAGTTT GAGGGGAAGA GAAAAGGAAA | 360 |
| GCTTGGGAGG TGAAGTGGGT ACCAGAGACA AGCAAGAAGA GCTGGTAGAA GTCATCTCAT | 420 |
| CTTAGGCTAC AATGAGGAAT TGAGACCTAG GAAGAAGGGA CACAGCAGGT AGAGAAACGT | 480 |
| GGCTTCTTGA CTCCAAGCC AGGAATTTGG GGAAAGGGGT TGGAGACCAT ACAAGGCAGA | 540 |
| GGGATGAGTG GGGAGAAGAA AGAAGGGAGA AAGGAAAGAT GGTGTACTCA CTCATTTGGG | 600 |
| ACTCAGGACT GAAGTGCCCA CTCACTTTTT TTTTTTTTTT TTTTGAGACA AACTTTCACT | 660 |
| TTTGTTGCCC AGGCTGGAGT GCAATGGCGC GATCTCGGCT CACTGCAACC TCCACCTCCC | 720 |
| GGGTTCAAGT GATTCTCCTG CCTCAGCCTC TAGCCAAGTA GCTGCGATTA CAGGCATGCG | 780 |
| CCACCACGCC CGGCTAATTT TTGTATTTTT AGTAGAGACG GGGTTTCGCC ATGTTGGTCA | 840 |
| GGCTGGTCTC GAACTCCTGA TCTCAGGTGA TCCAACCACC CTGGCCTCCC AAAGTGCTGG | 900 |
| GATTATAGGC GTGAGCCACA GCGCCTGGCC TGAAGCAGCC ACTCACTTTT ACAGACCCTA | 960 |
| AGACAATGAT TGCAAGCTGG TAGGATTGCT GTTTGGCCCA CCCAGCTGCG GTGTTGAGTT | 1020 |
| TGGGTGCGGT CTCCTGTGCT TTGCACCTGG CCCGCTTAAG GCATTTGTTA CCCGTAATGC | 1080 |
| TCCTGTAAGG CATCTGCGTT TGTGACATCG TTTTGGTCGC CAGGAAGGGA TTGGGGCTCT | 1140 |
| AAGCTTGAGC GGTTCATCCT TTTCATTTAT ACAG | 1174 |

(2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 193 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 2nd MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

| | |
|---|-----|
| GTGAGACACC CACCCGCTGC ACAGACCCAA TCTGGGAACC CAGCTCTGTG GATCTCCCCT | 60 |
| ACAGCCGTCC CTGAACACTG GTCCCGGGCG TCCCACCCGC CGCCACCGT CCCACCCCCT | 120 |
| CACCTTTTCT ACCCGGGTTC CCTAAGTTCC TGACCTAGGC GTCAGACTTC CTCACTATAC | 180 |
| TCTCCCACCC CAG | 193 |

(2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 131 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 3rd MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

| | |
|---|-----|
| GTGAGGGGGT CTCCCCGCCG AGACTTGGGG ATGGGGCGGG GCGCAGGGAA GGGAACCGTC | 60 |
| GCGCAGTGCC TGCCCCGGGG TTGGGCTGGC CCTACCGGGC GGGGCCGGCT CACTTGCCTC | 120 |
| TCCCTACGCA G | 131 |

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 89 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 4th MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

| | |
|---|----|
| GTGAGCGCGG ACTGGCCGAG AAGGGGCAAA GGAGCGGGGC GGACGGGGGC CAGAGACGTG | 60 |
| GCCCTCTCCT ACCCTCGTGT CCTTTTCAG | 89 |

(2) INFORMATION FOR SEQ ID NO: 43:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1400 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5th MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

| | |
|--|------|
| GTACCAGATC CTGGACACCC CCTACTCCCC GCTTTCCCAT CCCATGCTCC TCCCGGACTC | 60 |
| TATCGTGGAG CCAGAGACCC CATCCCAGCA AGCTCACTCA GGCCCCTGGC TGACAAACTC | 120 |
| ATTCACGCAC TGTTTGTTC TTTAACACCC ACTGTGAACC AGGCACCAGC CCCCACAAG | 180 |
| GATTCTGAAG CTGTAGGTCC TTGCCTCTAA GGAGCCCACA GCCAGTGGGG GAGGCTGACA | 240 |
| TGACAGACAC ATAGGAAGGA CATAGTAAAG ATGGTGGTCA CAGAGGAGGT GACACTTAAA | 300 |
| GCCTTCACTG GTAGAAAAGA AAAGGAGGTG TTCATTGCAG AGGAAACAGA ATGTGCAAAG | 360 |
| ACTCAGAATA TGGCCTATTT AGGGAATGGC TACATACACC ATGATTAGAG GAGGCCCAGT | 420 |
| AAAGGGAAGG GATGGTGAGA TGCCTGCTAG GTTCACTCAC TCACTTTTAT TTATTTATTT | 480 |
| ATTTTTTTGA CAGTCTCTCT GTCGCCCAGG CTGGAGTGCA GTGGTGTGAT CTTGGGTCAC | 540 |
| TGCAACTTCC GCCTCCCGGG TTCAAGGGAT TCTCCTGCCT CAGCTTCCTG AGTAGCTGGG | 600 |
| GTTACAGGTG TGTGCCACCA TGCCCAGCTA ATTTTTTTTT GTATTTTATAG TAGACAGGGT | 660 |
| TTCACCATGT TGGTCAGGCT GGTCTCAAAC TCCTGGCCTC AAGTGATCCG CCTGACTCAG | 720 |
| CCTACCAAAG TGCTGATTAC AAGTGTGAGC CACCGTGCCC AGCCACACTC ACTGATTCTT | 780 |
| TAATGCCAGC CACACAGCAC AAAGTTCAGA GAAATGCCTC CATCATAGCA TGTCAATATG | 840 |
| TTCATACTCT TAGGTTTCATG ATGTTCTTAA CATTAGGTTC ATAAGCAAAA TAAGAAAAAA | 900 |
| GAATAATAAA TAAAAGAAGT GGCATGTCAG GACCTCACCT GAAAAGCCAA ACACAGAATC | 960 |
| ATGAAGGTGA ATGCAGAGGT GACACCAACA CAAAGGTGTA TATATGGTTT CCTGTGGGGA | 1020 |

| | |
|--|------|
| GTATGTACGG AGGCAGCAGT GAGTGAGACT GCAAACGTCA GAAGGGCACG GGTCACCTGAG | 1080 |
| AGCCTAGTAT CCTAGTAAAG TGGGCTCTCT CCCTCTCTCT CCAGCTTGTC ATTGAAAACC | 1140 |
| AGTCCACCAA GCTTGTTGGT TCGCACAGCA AGAGTACATA GAGTTTGAAA TAATACATAG | 1200 |
| GATTTTAAGA GGGAGACACT GTCTCTAAAA AAAAAACAA CAGCAACAAC AAAAAGCAAC | 1260 |
| AACCATTACA ATTTTATGTT CCCTCAGCAT TCTCAGAGCT GAGGAATGGG AGAGGACTAT | 1320 |
| GGAACCCCC TTCATGTTCC GGCCTTCAGC CATGGCCCTG GATACATGCA CTCATCTGTC | 1380 |
| TTACAATGTC ATTCCCCCAG | 1400 |

(2) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1334 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 6th MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

| | |
|---|-----|
| GTCAGTTTGT TGGTCTGGCC ACTAATCTCT GTGGCCTAGT TCATAAAGAA TCACCCTTTG | 60 |
| GAGCTTCAGG TCTGAGGCTG GAGATGGGCT CCCTCCAGTG CAGGAGGGAT TGAAGCATGA | 120 |
| GCCAGCGCTC ATCTTGATAA TAACCATGAA GCTGACAGAC ACAGTTACCC GCAAACGGCT | 180 |
| GCCTACAGAT TGAAAACCAA GCAAAAACCG CCGGGCACGG TGGCTCACGC CTGTAATCCC | 240 |
| AGCACTTTGG GAGGCCAAGG CAGGTGGATC ACGAGGTCAA GAGATCAAGA CCATCCTGGC | 300 |

| | |
|---|------|
| CAACATGGTG AAACCCCATC TCTACTAAAA ATACGAAAAA ATAGCCAGGC GTGGTGGCGG | 360 |
| GTGCCTGTAA TCCCAGCTAC TCGGGAGGCT GAGGCAGGAG AATGGCATGA ACCCGGGAGG | 420 |
| CAGAAGTTGC AGTGAGCCGA GATCGTGCCA CTGCACTCCA GCCTGGGCAA CAGAGCGAGA | 480 |
| CTCTTGTCTC AAAAAAAAAA AAAAAAAAAAGA AAACCAAGCA AAAACCAAAA TGAGACAAAA | 540 |
| AAAACAAGAC CAAAAAATGG TGTTTGGAAG TTGTCAAGGT CAAGTCTGGA GAGCTAAACT | 600 |
| TTTTCTGAGA ACTGTTTATC TTTAATAAGC ATCAAATATT TTAACCTTGT AAATACTTTT | 660 |
| GTTGGAAATC GTTCTCTTCT TAGTCACTCT TGGGTCATTT TAAATCTCAC TTACTCTACT | 720 |
| AGACCTTTTA GGTTCCTGCT AGACTAGGTA GAACTCTGCC TTTGCATTTC TTGTGTCTGT | 780 |
| TTTGTATAGT TATCAATATT CATATTTATT TACAAGTTAT TCAGATCATT TTTTCTTTTC | 840 |
| TTTTTTTTTT TTTTTTTTTT TTTTACATCT TTAGTAGAGA CAGGGTTTCA CCATATTGGC | 900 |
| CAGGCTGCTC TCAAACCTCT GACCTTGTGA TCCACCAGCC TCGGCCTCCC AAAGTGCTGG | 960 |
| GATTCATTTT TTCTTTTTAA TTTGCTCTGG GCTTAAACTT GTGGCCCAGC ACTTTATGAT | 1020 |
| GGTACACAGA GTTAAGAGTG TAGACTCAGA CGGTCTTTCT TCTTTCCTTC TCTTCCTTCC | 1080 |
| TCCCTTCCCT CCCACCTTCC CTTCTCTCCT TCCTTTCTTT CTTCTCTCTT TGCTTCCTCA | 1140 |
| GGCCTCTTCC AGTTGCTCCA AAGCCCTGTA CTTTTTTTTT AGTTAACGTC TTATGGGAAG | 1200 |
| GGCCTGCACT TAGTGAAGAA GTGGTCTCAG AGTTGAGTTA CCTTGGCTTC TGGGAGGTGA | 1260 |
| AACTGTATCC CTATACCCTG AAGCTTTAAG GGGGTGCAAT GTAGATGAGA CCCCACATA | 1320 |
| GATCCTCTTC ACAG | 1334 |

(2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 512 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 7th MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

| | |
|---|-----|
| GTGGGCCTGG GGTGTGTGTG GACACAGTGG GTGCGGGGGA AAGAGGATGT AAGATGAGAT | 60 |
| GAGAAACAGG AGAAGAAAGA AATCAAGGCT GGGCTCTGTG GCTTACGCCT ATAATCCCAC | 120 |
| CACGTTGGGA GGCTGAGGTG GGAGAATGGT TTGAGCCCAG GAGTTCAAGA CAAGGCGGGG | 180 |
| CAACATAGTG TGACCCCATC TCTACCAAAA AAACCCCAAC AAAACCAAAA ATAGCCGGGC | 240 |
| ATGGTGGTAT GCGGCCTAGT CCCAGCTACT CAAGGAGGCT GAGGTGGGAA GATCGCTTGA | 300 |
| TTCCAGGAGT TTGAGACTGC AGTGAGCTAT GATCCCACCA CTGCCTACCA TCTTTAGGAT | 360 |
| ACATTTATTT ATTTATAAAA GAAATCAAGA GGCTGGATGG GGAATACAGG AGCTGGAGGG | 420 |
| TGGAGCCCTG AGGTGCTGGT TGTGAGCTGG CCTGGGACCC TTGTTTCCTG TCATGCCATG | 480 |
| AACCCACCCA CACTGTCCAC TGACCTCCCT AG | 512 |

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 114 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 8th MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

| | |
|--|-----|
| GTACAGCTTT GTCTGGTTTC CCCCAGCCA GTAGTCCCTT ATCCTCCCAT GTGTGTGCCA | 60 |
| GTGTCTGTCA TTGGTGGTCA CAGCCCGCCT CTCACATCTC CTTTTTCTCT CCAG | 114 |

(2) INFORMATION FOR SEQ ID NO: 47:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 617 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 9th MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

| | |
|--|-----|
| GTGAGTCTGC CCCTCCTCTT GGTCCCTGATG CCAGGAGACT CCTCAGCACC ATTCAGCCCC | 60 |
| AGGGCTGCTC AGGACCGCCT CTGCTCCCTC TCCTTTTCTG CAGAACAGAC CCCAACCCCA | 120 |
| ATATTAGAGA GGCAGATCAT GGTGGGGATT CCCCATTGT CCCAGAGGC TAATTGATTA | 180 |
| GAATGAAGCT TGAGAAATCT CCCAGCATCC CTCTCGCAA AGAATCCCC CCCCTTTTTT | 240 |
| TAAAGATAGG GTCTCACTCT GTTTGCCCCA GGCTGGGGTG TTGTGGCACG ATCATAGCTC | 300 |
| ACTGCAGCCT CGAACTCCTA GGCTCAGGCA ATCCTTTCAC CTTAGCTTCT CAAAGCACTG | 360 |
| GGACTGTAGG CATGAGCCAC TGTGCCTGGC CCCAAACGGC CCTTTTACTT GGCTTTTAGG | 420 |
| AAGCAAAAAC GGTGCTTATC TTACCCCTTC TCGTGTATCC ACCCTCATCC CTTGGCTGGC | 480 |
| CTCTTCTGGA GACTGAGGCA CTATGGGGCT GCCTGAGAAC TCGGGGCAGG GGTGGTGGAG | 540 |

TGCACTGAGG CAGGTGTTGA GGA ACTCTGC AGACCCCTCT TCCTTCCCAA AGCAGCCCTC 600
TCTGCTCTCC ATCGCAG 617

(2) INFORMATION FOR SEQ ID NO: 48:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 130 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 10th MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

GTATTACACT GACCCTTTCT TCAGGCACAA GCTTCCCCCA CCCTTGTGGA GTCACTTCAT 60
GCAAAGCGCA TGCAAATGAG CTGCTCCTGG GCCAGTTTTC TGATTAGCCT TTCCTGTTGT 120
GTACACACAG 130

(2) INFORMATION FOR SEQ ID NO: 49:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1401 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: Spans 3' part of 1st intron to beyond
end of 5th exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

| | |
|--|------|
| CAAAC TTTCA CTTTGT TGC CCAGGCTGGA GTGCAATGGC GCGATCTCGG CTCACTGCAA | 60 |
| CCTCCACCTC CCGGGTTCAA GTGATTCTCC TGCCTCAGCC TCTAGCCAAG TAGCTGCGAT | 120 |
| TACAGGCATG CGCCACCACG CCCGGCTAAT TTTTGTATTT TTAGTAGAGA CGGGGTTTCG | 180 |
| CCATGTTGGT CAGGCTGGTC TCGAACTCCT GATCTCAGGT GATCCAACCA CCCTGGCCTC | 240 |
| CCAAAGTGCT GGGATTATAG GCGTGAGCCA CAGCGCCTGG CCTGAAGCAG CCACTCACTT | 300 |
| TTACAGACCC TAAGACAATG ATTGCAAGCT GGTAGGATTG CTGTTTGGCC CACCCAGCTG | 360 |
| CGGTGTTGAG TTTGGGTGCG GTCTCCTGTG CTTTGCACCT GGCCCGCTTA AGGCATTTGT | 420 |
| TACCCGTAAT GCTCCTGTAA GGCATCTGCG TTTGTGACAT CGTTTTGGTC GCCAGGAAGG | 480 |
| GATTGGGGCT CTAAGCTTGA GCGGTTTCATC CTTTTCATTT ATACAGGGGA TGACCAGAGT | 540 |
| CATTGGCGCT ATGGAGGTGA GACACCCACC CGCTGCACAG ACCCAATCTG GGAACCCAGC | 600 |
| TCTGTGGATC TCCCCTACAG CCGTCCCTGA AACTGGTCC CGGGCGTCCC ACCCGCCGCC | 660 |
| CACCGTCCCA CCCCCTCACC TTTTCTACCC GGGTTCCTTA AGTTCCTGAC CTAGGCGTCA | 720 |
| GAATTCTCTA CTATACTCTC CCACCCAGG CGACCCGCCC TGGCCCCGGG TGTCCCCAGC | 780 |
| CTGCGCGGGC CGCTTCCAGT CCCCAGTGGA TATCCGCCCC CAGCTCGCCG CCTTCTGCCC | 840 |
| GGCCCTGCGC CCCCTGGAAC TCCTGGGCTT CCAGCTCCCG CCGCTCCCAG AACTGCGCCT | 900 |
| GCGCAACAAT GGCCACAGTG GTGAGGGGGT CTCCCCGCCG AGACTTGGGG ATGGGGCGGG | 960 |
| GCGCAGGGAA GGAACCGTC GCGCAGTGCC TGCCCCGGGG TTGGGCTGGC CCTACCGGGC | 1020 |
| GGGGCCGGCT CACTTGCTC TCCCTACGCA GTGCAACTGA CCCTGCCTCC TGGGCTAGAG | 1080 |
| ATGGCTCTGG GTCCCGGGCG GGAGTACCGG GCTCTGCAGC TGCATCTGCA CTGGGGGGCT | 1140 |

GCAGGTCGTC CGGGCTCGGA GCACACTGTG GAAGGCCACC GTTCCCTGC CGAGGTGAGC 1200
 GCGGACTGGC CGAGAAGGGG CAAAGGAGCG GGGCGGACGG GGGCCAGAGA CGTGGCCCTC 1260
 TCCTACCCTC GTGTCCTTTT CAGATCCACG TGGTTCACCT CAGCACCGCC TTTGCCAGAG 1320
 TTGACGAGGC CTTGGGGCGC CCGGGAGGCC TGGCCGTGTT GGCCGCCTTT CTGGAGGTAC 1380
 CAGATCCTGG ACACCCCTA C 1401

(2) INFORMATION FOR SEQ ID NO: 50:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 98 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

- (A) DESCRIPTION: Region of homology to collagen alpha 1 chain

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

| | | | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|----|----|----|
| Gln | Arg | Leu | Pro | Arg | Met | Gln | Glu | Asp | Ser | Pro | Leu | Gly | Gly | Gly | Ser | 1 | 5 | 10 | 15 |
| Ser | Gly | Glu | Asp | Asp | Pro | Leu | Gly | Glu | Glu | Asp | Leu | Pro | Ser | Glu | Glu | 20 | 25 | 30 | |
| Asp | Ser | Pro | Arg | Glu | Glu | Asp | Pro | Pro | Gly | Glu | Glu | Asp | Leu | Pro | Gly | 35 | 40 | 45 | |
| Glu | Glu | Asp | Leu | Pro | Gly | Glu | Glu | Asp | Leu | Pro | Glu | Val | Lys | Pro | Lys | 50 | 55 | 60 | |
| Ser | Glu | Glu | Glu | Gly | Ser | Leu | Lys | Leu | Glu | Asp | Leu | Pro | Thr | Val | Glu | 65 | 70 | 75 | 80 |
| Ala | Pro | Gly | Asp | Pro | Gln | Glu | Pro | Gln | Asn | Asn | Ala | His | Arg | Asp | Lys | 85 | 90 | 95 | |
| Glu | Gly | | | | | | | | | | | | | | | | | | |

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 256 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

- (A) DESCRIPTION: carbonic anhydrase domain

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Asp | Asp | Gln | Ser | His | Trp | Arg | Tyr | Gly | Gly | Asp | Pro | Pro | Trp | Pro | Arg |
| 1 | | | | 5 | | | | 10 | | | | | 15 | | |
| Val | Ser | Pro | Ala | Cys | Ala | Gly | Arg | Phe | Gln | Ser | Pro | Val | Asp | Ile | Arg |
| | | | 20 | | | | | 25 | | | | | 30 | | |
| Pro | Gln | Leu | Ala | Ala | Phe | Cys | Pro | Ala | Leu | Arg | Pro | Leu | Glu | Leu | Leu |
| | | | 35 | | | | | 40 | | | | | 45 | | |
| Gly | Phe | Gln | Leu | Pro | Pro | Leu | Pro | Glu | Leu | Arg | Leu | Arg | Asn | Asn | Gly |
| | 50 | | | | | 55 | | | | | 60 | | | | |
| His | Ser | Val | Gln | Leu | Thr | Leu | Pro | Pro | Gly | Leu | Glu | Met | Ala | Leu | Gly |
| 65 | | | | | 70 | | | | | 75 | | | | | 80 |
| Pro | Gly | Arg | Glu | Tyr | Arg | Ala | Leu | Gln | Leu | His | Leu | His | Trp | Gly | Ala |
| | | | | 85 | | | | | 90 | | | | | 95 | |
| Ala | Gly | Arg | Pro | Gly | Ser | Glu | His | Thr | Val | Glu | Gly | His | Arg | Phe | Pro |
| | | | 100 | | | | | 105 | | | | | 110 | | |
| Ala | Glu | Ile | His | Val | Val | His | Leu | Ser | Thr | Ala | Phe | Ala | Arg | Val | Asp |
| | | | 115 | | | | 120 | | | | | 125 | | | |
| Glu | Ala | Leu | Gly | Arg | Pro | Gly | Gly | Leu | Ala | Val | Leu | Ala | Ala | Phe | Leu |
| | 130 | | | | | 135 | | | | | 140 | | | | |
| Glu | Glu | Gly | Pro | Glu | Glu | Asn | Ser | Ala | Tyr | Glu | Gln | Leu | Leu | Ser | Arg |
| 145 | | | | | 150 | | | | | 155 | | | | | 160 |

| | | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|--|--|
| Leu | Glu | Glu | Ile | Ala | Glu | Glu | Gly | Ser | Glu | Thr | Gln | Val | Pro | Gly | Leu | | | |
| | | | | 165 | | | | | 170 | | | | | 175 | | | | |
| Asp | Ile | Ser | Ala | Leu | Leu | Pro | Ser | Asp | Phe | Ser | Arg | Tyr | Phe | Gln | Tyr | | | |
| | | | 180 | | | | | 185 | | | | | 190 | | | | | |
| Glu | Gly | Ser | Leu | Thr | Thr | Pro | Pro | Cys | Ala | Gln | Gly | Val | Ile | Trp | Thr | | | |
| | | 195 | | | | | 200 | | | | | 205 | | | | | | |
| Val | Phe | Asn | Gln | Thr | Val | Met | Leu | Ser | Ala | Lys | Gln | Leu | His | Thr | Leu | | | |
| | 210 | | | | | 215 | | | | | 220 | | | | | | | |
| Ser | Asp | Thr | Leu | Trp | Gly | Pro | Gly | Asp | Ser | Arg | Leu | Gln | Leu | Asn | Phe | | | |
| 225 | | | | | 230 | | | | | 235 | | | | | 240 | | | |
| Arg | Ala | Thr | Gln | Pro | Leu | Asn | Gly | Arg | Val | Ile | Glu | Ala | Ser | Phe | Pro | | | |
| | | | | 245 | | | | | 250 | | | | | 255 | | | | |

(2) INFORMATION FOR SEQ ID NO: 52:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (A) DESCRIPTION: transmembrane region

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

| | | | | | | | | | | | | | | | |
|-----------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Asp | Ile | Leu | Ala | Leu | Val | Phe | Gly | Leu | Leu | Phe | Ala | Val | Thr | Ser | Val |
| 1 | | | | 5 | | | | 10 | | | | | | 15 | |
| Ala Phe Leu Val | | | | | | | | | | | | | | | |
| 20 | | | | | | | | | | | | | | | |

(2) INFORMATION FOR SEQ ID NO: 53:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 amino acids
- (B) TYPE: amino acid

- (C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(A) DESCRIPTION: intracellular C-terminus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Arg | Arg | Gln | His | Arg | Arg | Gly | Thr | Lys | Gly | Gly | Val | Ser | Tyr | Arg |
| 1 | | | | 5 | | | | | 10 | | | | | 15 | |
| Pro | Ala | Glu | Val | Ala | Glu | Thr | Gly | Ala | | | | | | | |
| | | | 20 | | | | | 25 | | | | | | | |

(2) INFORMATION FOR SEQ ID NO: 54:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 170 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Arg | Ala | Leu | Gln | Leu | His | Leu | His | Trp | Gly | Ala | Ala | Gly | Arg | Pro | Gly |
| 1 | | | | 5 | | | | | 10 | | | | | 15 | |
| Ser | Glu | His | Thr | Val | Glu | Gly | His | Arg | Phe | Pro | Ala | Glu | Ile | His | Val |
| | | | 20 | | | | | 25 | | | | | 30 | | |
| Val | His | Leu | Ser | Thr | Ala | Phe | Ala | Arg | Val | Asp | Glu | Ala | Leu | Gly | Arg |
| | | 35 | | | | | 40 | | | | | 45 | | | |
| Pro | Gly | Gly | Leu | Ala | Val | Leu | Ala | Ala | Phe | Leu | Glu | Glu | Gly | Pro | Glu |
| | 50 | | | | | 55 | | | | | 60 | | | | |
| Glu | Asn | Ser | Ala | Tyr | Glu | Gln | Leu | Leu | Ser | Arg | Leu | Glu | Glu | Ile | Ala |
| 65 | | | | | 70 | | | | | 75 | | | | 80 | |
| Glu | Glu | Gly | Ser | Glu | Thr | Gln | Val | Pro | Gly | Leu | Asp | Ile | Ser | Ala | Leu |

| | | | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|--|--|--|
| | | | | 85 | | | | | 90 | | | | | 95 | | | | | |
| Leu | Pro | Ser | Asp | Phe | Ser | Arg | Tyr | Phe | Gln | Tyr | Glu | Gly | Ser | Leu | Thr | | | | |
| | | | 100 | | | | | 105 | | | | | 110 | | | | | | |
| Thr | Pro | Pro | Cys | Ala | Gln | Gly | Val | Ile | Trp | Thr | Val | Phe | Asn | Gln | Thr | | | | |
| | | 115 | | | | | 120 | | | | | 125 | | | | | | | |
| Val | Met | Leu | Ser | Ala | Lys | Gln | Leu | His | Thr | Leu | Ser | Asp | Thr | Leu | Trp | | | | |
| | 130 | | | | | 135 | | | | | | 140 | | | | | | | |
| Gly | Pro | Gly | Asp | Ser | Arg | Leu | Gln | Leu | Asn | Phe | Arg | Ala | Thr | Gln | Pro | | | | |
| 145 | | | | | 150 | | | | | 155 | | | | | 160 | | | | |
| Leu | Asn | Gly | Arg | Val | Ile | Glu | Ala | Ser | Phe | | | | | | | | | | |
| | | | | 165 | | | | | 170 | | | | | | | | | | |

(2) INFORMATION FOR SEQ ID NO: 55:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 470 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: RNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

| | |
|---|-----|
| CAUGGCCCCG AUAACCUUCU GCCUGUGCAC ACACCUGCCC CUCACUCCAC CCCCAUCCUA | 60 |
| GCUUUGGUU GGGGAGAGG GCACAGGGCC AGACAAACCU GUGAGACUUU GGCUCCAUCU | 120 |
| CUGCAAAGG GCGCUCUGUG AGUCAGCCUG CUCCCCUCCA GGCUUGCUC UCCCCACCC | 180 |
| AGCUCUCGUU UCCAAUGCAC GUACAGCCCC UACACACCGU GUGCUGGGAC ACCCCACAGU | 240 |
| CAGCCGCAUG GCUCCCCUGU GCCCCAGCCC CUGGCUCCCU CUGUUGAUCC CGGCCCCUGC | 300 |
| UCCAGGCCUC ACUGUGCAAC UGCUGCUGUC ACUGCUGCUU CUGGUGCCUG UCCAUCCCCA | 360 |
| GAGGUUGCCC CGGAUGCAGG AGGAUUCUCC CUUGGGAGGA GGCUCUUCUG GGGAAGAUGA | 420 |

(2) INFORMATION FOR SEQ ID NO: 56:

(i) SEQUENCE CHARACTERISTICS:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

This sequence is intentionally skipped.

(2) INFORMATION FOR SEQ ID NO: 57:

(i) SEQUENCE CHARACTERISTICS:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

This sequence is intentionally skipped.

(2) INFORMATION FOR SEQ ID NO: 58:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 904 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

| | |
|---|-----|
| GCTGGTCTCG AACTCCTGGA CTCAAGCAAT CCACCCACCT CAGCCTCCCA AAATGAGGGA | 60 |
| CCGTGTCTTA TTCATTTCCA TGTCCCTAGT CCATAGCCCA GTGCTGGACC TATGGTAGTA | 120 |
| CTAAATAAAT ATTTGTTGAA TGCAATAGTA AATAGCATTT CAGGGAGCAA GAACTAGATT | 180 |
| AACAAAGGTG GTAAAAGGTT TGGAGAAAAA AATAATAGTT TAATTTGGCT AGAGTATGAG | 240 |

| | |
|--|-----|
| GGAGAGTAGT AGGAGACAAG ATGGAAAGGT CTCTTGGGCA AGGTTTTGAA GGAAGTTGGA | 300 |
| AGTCAGAAAGT ACACAATGTG CATATCGTGG CAGGCAGTGG GGAGCCAATG AAGGCTTTTG | 360 |
| AGCAGGAGAG TAATGTGTTG AAAAATAAAT ATAGGTTAAA CCTATCAGAG CCCCTCTGAC | 420 |
| ACATACACTT GCTTTTCATT CAAGCTCAAG TTTGTCTCCC ACATACCCAT TACTTAACTC | 480 |
| ACCCTCGGGC TCCCCTAGCA GCCTGCCCTA CCTCTTTACC TGCTTCCTGG TGGAGTCAGG | 540 |
| GATGTATACA TGAGCTGCTT TCCCTCTCAG CCAGAGGACA TGGGGGGCCC CAGCTCCCCT | 600 |
| GCCTTTCCCC TTCTGTGCCT GGAGCTGGGA AGCAGGCCAG GGTTAGCTGA GGCTGGCTGG | 660 |
| CAAGCAGCTG GGTGGTGCCA GGGAGAGCCT GCATAGTGCC AGGTGGTGCC TTGGGTTCCA | 720 |
| AGCTAGTCCA TGGCCCCGAT AACCTTCTGC CTGTGCACAC ACCTGCCCCCT CACTCCACCC | 780 |
| CCATCCTAGC TTTGGTATGG GGGAGAGGGC ACAGGGCCAG ACAAACCTGT GAGACTTTGG | 840 |
| CTCCATCTCT GCAAAAGGGC GCTCTGTGAG TCAGCCTGCT CCCCTCCAGG CTTGCTCCTC | 900 |
| CCCC | 904 |

(2) INFORMATION FOR SEQ ID NO: 59:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 292 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

| | |
|--|-----|
| TTTTTTTGGAG ACGGAGTCTT GCATCTGTCA TGCCCAGGCT GGAGTAGCAG TGGTGCCATC | 60 |
| TCGGCTCACT GCAAGCTCCA CCTCCCGAGT TCACGCCATT TTCCTGCCTC AGCCTCCCGA | 120 |

| | |
|---|-----|
| GTAGCTGGGA CTACAGGCGC CCGCCACCAT GCCCGGCTAA TTTTTTGTAT TTTTGGTAGA | 180 |
| GACGGGGTTT CACCGTGTTA GCCAGAATGG TCTCGATCTC CTGACTTCGT GATCCACCCG | 240 |
| CCTCGGCCTC CCAAAGTTCT GGGATTACAG GTGTGAGCCA CCGCACCTGG CC | 292 |

(2) INFORMATION FOR SEQ ID NO: 60:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 262 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

| | |
|--|-----|
| TTCTTTTTTG AGACAGGGTC TTGCTCTGTC ACCCAGGCCA GAGTGCAATG GTACAGTCTC | 60 |
| AGCTCACTGC AGCCTCAACC GCCTCGGCTC AAACCATCAT CCCATTTTCAG CCTCCTGAGT | 120 |
| AGCTGGGACT ACAGGCACAT GCCATTACAC CTGGCTAATT TTTTGTATT TCTAGTAGAG | 180 |
| ACAGGGTTTG GCCATGTTGC CCGGGCTGGT CTCGAACTCC TGGACTCAAG CAATCCACCC | 240 |
| ACCTCAGCCT CCCAAAATGA GG | 262 |

(2) INFORMATION FOR SEQ ID NO: 61:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 294 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

| | |
|--|-----|
| TTTTTTTTTTG AGACAAACTT TCACTTTTGT TGCCCAGGCT GGAGTGCAAT GGCGCGATCT | 60 |
| CGGCTCACTG CAACCTCCAC CTCCCGGGTT CAAGTGATTC TCCTGCCTCA GCCTCTAGCC | 120 |
| AAGTAGCTGC GATTACAGGC ATGCGCCACC ACGCCCGGCT AATTTTTGTA TTTTGTAGTAG | 180 |
| AGACGGGGTT TCGCCATGTT GGTCAGGCTG GTCTCGAACT CCTGATCTCA GGTGATCCAA | 240 |
| CCACCCTGGC CTCCCAAAGT GCTGGGATTA TAGGCGTGAG CCACAGCGCC TGGC | 294 |

(2) INFORMATION FOR SEQ ID NO: 62:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 276 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

| | |
|---|-----|
| TGACAGTCTC TCTGTCGCCC AGGCTGGAGT GCAGTGGTGT GATCTTGGGT CACTGCAACT | 60 |
| TCCGCCTCCC GGGTTCAAGG GATTCTCCTG CCTCAGCTTC CTGAGTAGCT GGGGTTACAG | 120 |
| GTGTGTGCCA CCATGCCCAG CTAATTTTTT TTTGTATTTT TAGTAGACAG GGTTTCACCA | 180 |
| TGTTGGTCAG GCTGGTCTCA AACTCCTGGC CTCAAGTGAT CCGCCTGACT CAGCCTACCA | 240 |
| AAGTGCTGAT TACAAGTGTG AGCCACCGTG CCCAGC | 276 |

(2) INFORMATION FOR SEQ ID NO: 63:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 289 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

| | |
|---|-----|
| CGCCGGGCAC GGTGGCTCAC GCCTGTAATC CCAGCACTTT GGGAGGCCAA GGCAGGTGGA | 60 |
| TCACGAGGTC AAGAGATCAA GACCATCCTG GCCAACATGG TGAAACCCCA TCTCTACTAA | 120 |
| AAATACGAAA AAATAGCCAG GCGTGGTGGC GGGTGCCTGT AATCCCAGCT ACTCGGGAGG | 180 |
| CTGAGGCAGG AGAATGGCAT GAACCCGGGA GGCAGAAGTT GCAGTGAGCC GAGATCGTGC | 240 |
| CACTGCACTC CAGCCTGGGC AACAGAGCGA GACTCTTGTC TCAAAAAAA | 289 |

(2) INFORMATION FOR SEQ ID NO: 64:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 298 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

| | |
|---|-----|
| AGGCTGGGCT CTGTGGCTTA CGCCTATAAT CCCACCACGT TGGGAGGCTG AGGTGGGAGA | 60 |
| ATGGTTTGAG CCCAGGAGTT CAAGACAAGG CGGGGCAACA TAGTGTGACC CCATCTCTAC | 120 |
| CAAAAAAACC CCAACAAAAC CAAAAATAGC CGGGCATGGT GGTATGCGGC CTAGTCCCAG | 180 |
| CTACTCAAGG AGGCTGAGGT GGGAAGATCG CTTGATTCCA GGAGTTTGAG ACTGCAGTGA | 240 |
| GCTATGATCC CACCACTGCC TACCATCTTT AGGATACATT TATTTATTTA TAAAAGAA | 298 |

(2) INFORMATION FOR SEQ ID NO: 65:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 105 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

| | |
|--|-----|
| TTTTTTACAT CTTTAGTAGA GACAGGGTTT CACCATATTG GCCAGGCTGC TCTCAAATC | 60 |
| CTGACCTTGT GATCCACCAG CCTCGGCCTC CCAAAGTGCT GGGAT | 105 |

(2) INFORMATION FOR SEQ ID NO: 66:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 83 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

CCTCGAACTC CTAGGCTCAG GCAATCCTTT CACCTTAGCT TCTCAAAGCA CTGGGACTGT 60

AGGCATGAGC CACTGTGCCT GGC 83

(2) INFORMATION FOR SEQ ID NO: 67:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

AGAAGGTAAG T 11

(2) INFORMATION FOR SEQ ID NO: 68:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

TGGAGGTGAG A 11

(2) INFORMATION FOR SEQ ID NO: 69:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 11 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

CAGTCGTGAG G

11

(2) INFORMATION FOR SEQ ID NO: 70:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 11 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

CCGAGGTGAG C

11

(2) INFORMATION FOR SEQ ID NO: 71:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 11 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

TGGAGGTACC A

11

(2) INFORMATION FOR SEQ ID NO: 72:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

GGAAGGTCAG T

11

(2) INFORMATION FOR SEQ ID NO: 73:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

AGCAGGTGGG C

11

(2) INFORMATION FOR SEQ ID NO: 74:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

GCCAGGTACA G

11

(2) INFORMATION FOR SEQ ID NO: 75:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

TGCTGGTGAG T

11

(2) INFORMATION FOR SEQ ID NO: 76:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

ATACAGGGGAT

11

(2) INFORMATION FOR SEQ ID NO: 77:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

ATACAGGGGA T

11

(2) INFORMATION FOR SEQ ID NO: 78:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

CCCCAGGCGA C

11

(2) INFORMATION FOR SEQ ID NO: 79:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

ACGCAGTGCA A

11

(2) INFORMATION FOR SEQ ID NO: 80:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

TTTCAGATCC A

11

(2) INFORMATION FOR SEQ ID NO: 81:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

CCCCAGGAGG G

11

(2) INFORMATION FOR SEQ ID NO: 82:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

TCACAGGCTC A

11

(2) INFORMATION FOR SEQ ID NO: 83:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

CCCTAGCTCC A

11

(2) INFORMATION FOR SEQ ID NO: 84:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

CTCCAGTCCA G

11

(2) INFORMATION FOR SEQ ID NO: 85:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 12 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

TCGCAGGTGA CA

12

(2) INFORMATION FOR SEQ ID NO: 86:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

ACACAGAAGG G